数理解析研究所講究録
第 1620 巻 2009 年 175-196

Some problems of algorithmic randomness on product space

統計数理研究所　高橋 勇人 (Hayato Takahashi)
The Institute of Statistical Mathematics.

1 Introduction

Intuitively, a sequence is random if it is not covered by a large class of null sets . A definition of random sequences depends on a class of null sets.

Martin-Löf randomness [12] is one of the definitions of randomness and defined by sequences that are not covered by null sets with effective manner. It is known that Martin-Löf random sequences satisfy many laws of probability one, for example ergodic theorem, martingale convergence theorem, and so on. In this paper, we study Martin-Löf random sequences with respect to a probability on product space $\Omega \times \Omega$, where Ω is the set of infinite binary sequences. In particular, we investigate the following problems:

1. randomness and monotone complexity on product space (Levin-Schnorr theorem for product space)
2. conditional probability and Fubini's theorem for individual random sequences.
3. section of random set vs. relativized randomness.
4. decomposition of complexity and independence of individual random sequences.
5. Bayesian statistics for individual random sequences.

The above problems are property of product space. Besides above problems, we show classification of random set by likelihood ratio test, which is necessary for 4 and 5.

2 Randomness and complexity

First we introduce Martin-Löf randomness on Ω. Let S be the set of finite binary strings. Let Ω be the set of infinite binary sequences with product topology. For $x \in S$, let $\Delta(x) := \{x\omega : \omega \in \Omega\}$, where $x\omega$ is the concatenation of x and ω. Let $\mathcal{B}$ be the σ-algebra generated by $\{\Delta(x)\}_{x\in S}$. Let $(P, \mathcal{B}, \Omega)$ be a probability space. We write $P(x) := P(\Delta(x))$ for $x \in S$, then we have $P(x) = P(x0) + P(x1)$ for all x. Let $\mathbb{N}$, $\mathbb{Q}$, and $\mathbb{R}$ be the set of natural numbers, rational numbers, and real numbers, respectively. P is called computable if there exists a computable function $p : S \times \mathbb{N} \to \mathbb{Q}$ such that $\forall x \in S \forall k \in \mathbb{N}\ |P(x) - p(x,k)| < 1/k$. A set $A \subset S$ is called recursively enumerable (r.e.) if there is a computable function $f : \mathbb{N} \to S$ such that $f(\mathbb{N}) = A$. For $A \subset S$, let $\tilde{A} := \cup_{x\in A}\Delta(x)$. A set $U \subset \mathbb{N} \times S$ is called (Martin-Löf) test with respect to P if 1) U is r.e., 2) $U_{n+1} \subset U_n$ for all n, where $U_n = \{x : (n,x) \in U\}$, and 3) $P(\tilde{U}_n) < 2^{-n}$. In the following, if P is obvious from the context, we say that U is a test. A test U is called universal if for any other test V, there is a constant c such that $\forall n\ V_{n+c} \subset U_n$.

Theorem 2.1 (Martin-Löf[12]) *If P is a computable probability, a universal test U exists.*

In [12], the set $(\cap_{n=1}^{\infty}\tilde{U}_n)^c$ (complement of the limit of universal test) is defined to be random sequences with respect to P, where U is a universal test. We write $\mathcal{R}^P := (\cap_{n=1}^{\infty}\tilde{U}_n)^c$. Note that for two universal tests U and V, $\cap_{n=1}^{\infty}\tilde{U}_n = \cap_{n=1}^{\infty}\tilde{V}_n$ and hence $\mathcal{R}^P$ does not depend on the choice of a universal test.

For $\mathbf{x} = (x^1, \cdots, x^k) \in S^k$, let $\Delta(\mathbf{x}) := \Delta(x^1) \times \cdots \times \Delta(x^k)$. Let P be a computable probability on $(\mathcal{B}_{\Omega^k}, \Omega^k)$, where $\mathcal{B}_{\Omega^k}$ is the Borel-σ-algebra generated by $\{\Delta(\mathbf{x})\}_{\mathbf{x}\in S^k}$. The computability of P is defined in the same way. Since there is a bijection $f : S \to S^k$ such that f and f^{-1} are computable, we can define a Martin-Löf test and a universal Martin-Löf test with respect to a computable probability on Ω^k in the same way. As in [12], we can show that a universal test U exists for a computable probability on Ω^k. Let $\mathcal{R}^P := (\cap_{n=1}^{\infty}\tilde{U}_n)^c \subset \Omega^k$. We call $\mathcal{R}^P$ the set of random sequences (or points) with respect to P.

Remark 1 We see that there is a bijection $g : S \to \cup_{k<\infty}S^k$ such that g and g^{-1} are computable. Hence, we can define a universal test with respect to a computable probability on $(\mathcal{B}_{\Omega^\infty}, \Omega^\infty)$ in the same way. In this paper, we primarily study the random points of computable probabilities on the finite dimensional product space Ω^k with product topology. For algorithmic randomness on other separable metric spaces including Ω^∞, see [9].

2.1 Complexity

Next, we introduce monotone complexity and we characterize $\mathcal{R}^P$ by it. In the following discussion, we generalize the monotone complexity defined in [10, 19] in a natural way. Let $|s|$ be the length of $s \in S$ and $\bar{s} := 1^{|s|}0s$. $|\lambda| = 0$, where λ is the empty word, and $|x^\infty| = \infty$ for $x^\infty \in \Omega$. For $\mathbf{s} = (s^1, \ldots, s^k) \in (S \cup \Omega)^k$, set

$$|\mathbf{s}| := |s^1| + \cdots + |s^k|.$$

We write $x \sqsubseteq y$ for $x, y \in S \cup \Omega$, if x is a prefix of y. For $x^\infty \in \Omega$, set $\Delta(x^\infty) := \{x^\infty\}$, and for $\mathbf{x} = (x^1, \cdots, x^k) \in (S \cup \Omega)^k$, set $\Delta(\mathbf{x}) := \Delta(x^1) \times \cdots \times \Delta(x^k)$. For $\mathbf{y} = (y^1, \cdots, y^k) \in (S \cup \Omega)^k$, we write $\mathbf{x} \sqsubseteq \mathbf{y}$ if $x^1 \sqsubseteq y^1, \cdots, x^k \sqsubseteq y^k$, i.e., $\mathbf{x} \sqsubseteq \mathbf{y} \Leftrightarrow \Delta(\mathbf{x}) \supset \Delta(\mathbf{y})$. $\mathbf{x}$ and $\mathbf{y}$ are called comparable if $\mathbf{x} \sqsubseteq \mathbf{y}$ or $\mathbf{y} \sqsubseteq \mathbf{x}$. Let $A \subset S^k$. $\mathbf{x} \in (S \cup \Omega)^k$ is called least upper bound of A if $\forall \mathbf{y} \in A$, $\mathbf{y} \sqsubseteq \mathbf{x}$ and if $\forall \mathbf{y} \in A$, $\mathbf{y} \sqsubseteq \mathbf{z}$ then $\mathbf{x} \sqsubseteq \mathbf{z}$. The least upper bound of A is denoted by $\sup A$. The $\sup A$ exists iff $\cap_{\mathbf{x} \in A} \Delta(\mathbf{x}) \neq \emptyset$. Note that if $\Delta(\mathbf{x}) \cap \Delta(\mathbf{y}) \neq \emptyset$ then there is $\mathbf{z}$ such that $\Delta(\mathbf{x}) \cap \Delta(\mathbf{y}) = \Delta(\mathbf{z})$. Thus if $\cap_{\mathbf{x} \in A} \Delta(\mathbf{x}) \neq \emptyset$, there is $\mathbf{y} \in (S \cup \Omega)^k$ such that $\cap_{\mathbf{x} \in A} \Delta(\mathbf{x}) = \Delta(\mathbf{y})$ and $\sup A = \mathbf{y}$. For example, $\sup\{(\lambda, 0), (0, \lambda)\} = (0, 0)$ and $\sup\{x | x \sqsubset x^\infty\} = x^\infty$.

In the following, when k is clear from the context, we use bold-faced symbols such as 1) $\mathbf{x}^\infty$, $\mathbf{y}^\infty$ to denote an element of Ω^k, 2) $\mathbf{x}$, $\mathbf{y}$, $\mathbf{s}$ to denote an element of S^k or $(S \cup \Omega)^k$ (we will specify which space we consider), and 3) $\boldsymbol{\lambda}$ to denote $(\lambda, \cdots, \lambda) \in S^k$ for $k \geq 1$, and $\Delta(\boldsymbol{\lambda}) = \Omega^k$. Further, we write $P(\mathbf{x})$ for $P(\Delta(\mathbf{x}))$.

Let $F \subset S^j \times S^k$ and $F_{\mathbf{s}} := \{\mathbf{x} | (\mathbf{s}, \mathbf{x}) \in F\}$. Assume that:
a0) $\forall \mathbf{s} \in S^j$, $\boldsymbol{\lambda} \in F_{\mathbf{s}}$.
a1) $\forall \mathbf{s} \in S^j$, $\sup \cup_{\mathbf{s}' \sqsubseteq \mathbf{s}} F_{\mathbf{s}'}$ exists, i.e., $\cap_{\mathbf{x} \in \cup_{\mathbf{s}' \sqsubseteq \mathbf{s}} F_{\mathbf{s}'}} \Delta(\mathbf{x}) \neq \emptyset$.
Set

$$f(\mathbf{s}) := \sup \bigcup_{\mathbf{s}' \sqsubseteq \mathbf{s},\ \mathbf{s}' \in S^j} F_{\mathbf{s}'} \text{ for } \mathbf{s} \in (S \cup \Omega)^j. \tag{1}$$

We see that $f : (S \cup \Omega)^j \to (S \cup \Omega)^k$ and f is monotonically increasing, i.e., $\mathbf{s}' \sqsubseteq \mathbf{s} \Rightarrow f(\mathbf{s}') \sqsubseteq f(\mathbf{s})$.

Conversely, let $f : (S \cup \Omega)^j \to (S \cup \Omega)^k$ be a monotonically increasing function, and set

$$F := \{(\mathbf{s}, \mathbf{x}) \in S^j \times S^k | \mathbf{x} \sqsubseteq f(\mathbf{s})\}.$$

Then $\sup F_{\mathbf{s}} = f(\mathbf{s})$, F satisfies a0 and a1, and the function defined by F coincides with f.

Now assume that
a2) F is a r.e. set.
Then the function f defined by (1) is called *computable monotone function.*

The monotone complexity with respect to computable monotone function $f : (S \cup \Omega)^{k+j} \to (S \cup \Omega)^k$ is defined as follows:

$$Km_f(\mathbf{x}|\mathbf{y}) := \min\{|\mathbf{p}| | \mathbf{x} \sqsubseteq f(\mathbf{p}, \mathbf{y})\},$$

where $\mathbf{p} \in (S \cup \Omega)^k$, $\mathbf{y} \in (S \cup \Omega)^j$, and $\mathbf{x} \in (S \cup \Omega)^k$. If there is no $\mathbf{p}$ such that $\mathbf{x} \sqsubseteq f(\mathbf{p}, \mathbf{y})$, then $Km_f(\mathbf{x}|\mathbf{y}) := \infty$. A $\mathbf{p}$ whose length attains $Km_f(\mathbf{x}|\mathbf{y})$ is called optimal code for $Km_f(\mathbf{x}|\mathbf{y})$. For each fixed dimension k, j, a computable monotone function $u : (S \cup \Omega)^{k+j} \to (S \cup \Omega)^k$ is called *optimal* if for any computable monotone function $f : (S \cup \Omega)^{k+j} \to (S \cup \Omega)^k$, there is a constant c such that $Km_u(\mathbf{x}|\mathbf{y}) \leq Km_f(\mathbf{x}|\mathbf{y}) + c$ for all $\mathbf{x} \in (S \cup \Omega)^k$, $\mathbf{y} \in (S \cup \Omega)^j$. We can construct an optimal function in the following manner. First, observe that there is a r.e. set $F' \subset \mathbb{N} \times S^{k+j} \times S^j$ such that 1) $F_i = \{(\mathbf{s}, \mathbf{x}) | (i, \mathbf{s}, \mathbf{x}) \in F'\}$ satisfies conditions a0–a2 and is defined for all $i \in \mathbb{N}$, and 2) for each F that satisfies conditions a0–a2, there is i such that $F = F_i$. Next, set $F^u := \{(c(i, \mathbf{s}), \mathbf{x}) | (i, \mathbf{s}, \mathbf{x}) \in F'\}$, where $c(i, \mathbf{s}) = (\bar{i}s^1, s^2, \cdots, s^{k+j})$ for $\mathbf{s} = (s^1, s^2, \cdots, s^{k+j})$, and computable monotone function $u : (S \cup \Omega)^{k+j} \to (S \cup \Omega)^k$ is defined by F^u via (1). In such a case, we see that u is optimal. In the following discussion, we fix an optimal function u for each dimension and let $Km(\mathbf{x}|\mathbf{y}) := Km_u(\mathbf{x}|\mathbf{y})$ and $Km(\mathbf{x}) := Km_u(\mathbf{x})$.

By definition, we have

Proposition 2.1 *1) Monotonicity:* $\mathbf{x} \sqsubseteq \mathbf{z} \Rightarrow Km(\mathbf{x}|\mathbf{y}) \leq Km(\mathbf{z}|\mathbf{y})$, *and* $\mathbf{y} \sqsubseteq \mathbf{z} \Rightarrow Km(\mathbf{x}|\mathbf{y}) \geq Km(\mathbf{x}|\mathbf{z})$.
2) Kraft inequality: $\forall \mathbf{y}$, $\sum_{\mathbf{x} \in \mathcal{A}} 2^{-Km(\mathbf{x}|\mathbf{y})} \leq 1$ *for prefix-free set* $\mathcal{A} \subset (S \cup \Omega)^k$, *where* $\mathcal{A}$ *is called prefix-free if* $\Delta(\mathbf{x}) \cap \Delta(\mathbf{y}) = \emptyset$ *for* $\mathbf{x}, \mathbf{y} \in \mathcal{A}, \mathbf{x} \neq \mathbf{y}$.
3) Conditional sub-additivity: $\exists c\ \forall \mathbf{x} \in S^k, \mathbf{y} \in S^j$, $Km(\mathbf{x}, \mathbf{y}) \leq Km(\mathbf{x}|\mathbf{y}) + Km(\mathbf{y}) + c$.

Theorem 2.2 (Levin-Schnorr[10, 15, 16]) *Let P be a computable probability on Ω. Then, $x^\infty \in \mathcal{R}^P$ iff* $\sup_{x \sqsubset x^\infty} - \log P(x) - Km(x) < \infty$.

Next we show a weak form of Levin-Schnorr theorem for product space. Before proving the theorem, we need conditions. Let $\mathcal{A} \subset S^k$.

Condition 1) if $\mathbf{x}, \mathbf{y} \in \mathcal{A}$ then, $\mathbf{x}$ and $\mathbf{y}$ are comparable or $\Delta(\mathbf{x}) \cap \Delta(\mathbf{y}) = \emptyset$.

Condition 2) there is a recursive function $f : S^k \times \mathbb{N} \to \mathcal{A}$ such that for any $\mathbf{x} \in S^k$, $\Delta(\mathbf{x}) = \widetilde{f(\mathbf{x}, \mathbb{N})}$ and $f(\mathbf{x}, \mathbb{N})$ is prefix-free.

For example, S satisfies conditions 1 and 2. $\{(x, y) \| x| = |y|\}$ satisfies conditions 1 and 2. $\{(x, \lambda) | x \in S\}$ satisfy condition 1 but it does not satisfy 2. $\{(x, y) \| x| = |y| + 1 \text{ or } |x| = |y| - 1\}$ satisfies 2 but it does not satisfy 1; in particular, S^2 itself satisfies 2 but it does not satisfy 1.

Lemma 2.1 *a) If $\mathcal{A}$ satisfies condition 1 then for any $B \subset \mathcal{A}$ there is a $C \subset B$ such that C is prefix-free and $\tilde{B} = \tilde{C}$.*
b) If a r.e. set $\mathcal{A}$ satisfies condition 2 then for any r.e. $B \subset S^k$, there is a r.e. $C \subset \mathcal{A}$ such that C is prefix-free and $\tilde{B} = \tilde{C}$.

Let $\mathcal{A}(\mathbf{x}^\infty) := \{\mathbf{x} \in S^k | \mathbf{x} \in \mathcal{A},\ \mathbf{x} \sqsubset \mathbf{x}^\infty\}$.

Theorem 2.3 (Levin-Schnorr theorem on product space) *Let P be a computable probability on Ω^k. If a r.e. set $\mathcal{A} \subset S^k$ satisfies conditions 1 and 2, then $\mathbf{x}^\infty \in \mathcal{R}^P$ iff $\sup_{\mathbf{x} \in \mathcal{A}(\mathbf{x}^\infty)} - \log P(\mathbf{x}) - Km(\mathbf{x}) < \infty$.*

Proof) If $\mathbf{x}^\infty \notin \mathcal{R}^P$, then for each n, there is a r.e. set S_n such that $\mathbf{x}^\infty \in \tilde{S}_n$ and $P(\tilde{S}_n) < 2^{-n}$. Since S_n is a r.e. set, by Lemma 2.1 b, we can construct a r.e. prefix-free set S'_n such that $S'_n \subset \mathcal{A}$ and $\tilde{S}'_n = \tilde{S}_n$. Let P' be a measure such that $P'(\mathbf{x}) = P(\mathbf{x})2^n$ for $\mathbf{x} \in S'_n$ and 0 otherwise; then, we have $\sum_{\mathbf{x} \in S'_n} P'(\mathbf{x}) < 1$. Since S'_n is a r.e. set, by applying Shannon-Fano-Elias coding to P', we see that there is a sequence $\{\mathbf{x}(n)\}$ of prefix of $\mathbf{x}^\infty$ such that $\forall n\ \mathbf{x}(n) \in \mathcal{A}$ and $\exists c_1, c_2 > 0 \forall n\ Km(\mathbf{x}(n)) \le -\log P(\mathbf{x}(n)) - n + K(n) + c_1 \le -\log P(\mathbf{x}(n)) - n + 2\log n + c_2$, where K is the prefix complexity.

Conversely, let

$$U_n := \{\mathbf{x} | \mathbf{x} \in \mathcal{A}, Km(\mathbf{x}) < -\log P(\mathbf{x}) - n\}.$$

By Lemma 2.1 a, there is a prefix-free set $U'_n \subset U_n$ such that $\tilde{U}'_n = \tilde{U}_n$, and hence $P(\tilde{U}_n) = P(\tilde{U}'_n) < \sum_{\mathbf{x} \in U'_n} 2^{-Km(\mathbf{x})-n} \le 2^{-n}$, where the last inequality follows from Proposition 2.1 2. Since U_n is a r.e. set, $\{U_n\}$ is a test and $\cap_n \tilde{U}_n \subset (\mathcal{R}^P)^c$. ■

The author do not know whether the right-hand-side of 2) of Theorem 2.3 is replaced with $\sup_{\mathbf{x} \sqsubset \mathbf{x}^\infty} - \log P(\mathbf{x}) - Km(\mathbf{x}) < \infty$ for $k \ge 2$. Recall that S^k ($k \ge 2$) itself does not satisfy condition 1.

In order to show a coding theorem, we introduce a class of partition. Let $f_i : \mathbb{N} \to \mathbb{N} \cup \{0\}$, $1 \le i \le k$ be monotonically increasing total recursive functions, and $f := (f_1, \ldots, f_{k-1})$. Then, set

$$\mathcal{A}_n^f := \{(x^1, \ldots, x^k) \in S^k | f(n) = (|x^1|, \ldots, |x^k|)\},$$

and $\mathcal{A}^f := \cup_n \mathcal{A}_n^f$. f is called *partition function.*

Lemma 2.2 *If f is unbounded then $\mathcal{A}^f$ satisfies conditions 1 and 2.*

If P is a computable probability on Ω, then by applying arithmetic coding, we have:

$$\sup_{x \in S} Km(x) + \log P(x) < \infty. \tag{2}$$

For more information on Shannon-Fano-Elias coding and arithmetic coding, see [7]. Further, see [19] for the proof of Theorem 2.2 and (2). If P is a computable probability on Ω^k, for $k \geq 2$, then the situation is different and it is not known whether (2) holds in the case of multiple dimensions. However if we restrict the domain of $\mathbf{x}$ to A^f, we have

Lemma 2.3 *Let P be a computable probability on Ω^k. Then, for any k-dimensional partition function f,*

$$\sup_{\mathbf{x} \in A^f} Km(\mathbf{x}) + \log P(\mathbf{x}) < \infty.$$

Thus, by Theorem 2.3, we have:

Corollary 2.1 *Let P be a computable probability on Ω^k. Then, for any k-dimensional unbounded partition function f,*

$$\mathbf{x}^\infty \in \mathcal{R}^P \Leftrightarrow \sup_{\mathbf{x} \in A^f(\mathbf{x}^\infty)} |\log P(\mathbf{x}) + Km(\mathbf{x})| < \infty.$$

In [6], a conditional complexity K_* that is monotone with the conditional argument is defined.

3 Martingale and conditional probability

Let P be a computable probability on Ω. Let $S_n := \{s||s| = n\}$ for $n \in \mathbb{N}$. Let $\mathcal{F}_n$ be the algebra generated by $\{\Delta(x)|x \in S_n\}$ and $\mathcal{F}_\infty := \sigma(\cup_n \mathcal{F}_n)$. Let $X_n : \Omega \to \mathbb{R}$ be a measurable function with respect to $\mathcal{F}_n$, i.e., X_n takes a constant value on $\Delta(x)$ for $|x| = n$. Let $X_n(x) := X_n(x^\infty)$ for $x^\infty \in \Delta(x)$

and $x \in S_n$. $\{X_n\}_{n\in\mathbb{N}}$ is called 1) submartingale if $\forall n,\ E(X_n|\mathcal{F}_{n-1}) \geq X_{n-1},\ P-a.s.$, and 2) martingale if $\forall n,\ E(X_n|\mathcal{F}_{n-1}) = X_{n-1},\ P-a.s.$ Let

$$D := \{x \in S|P(x) > 0\}.$$

We say that a submartingale $\{X_n\}$ is computable if there is a computable function $A : \mathbb{N}\times D\times\mathbb{N} \to \mathbb{Q}$ such that $\forall n\forall x \in S_n\cap D\forall k, |A(n,x,k)-X_n(x)| < 1/k$. In the above definition, X_n need not be computable on S_n. We require that X_n is computable on $S_n \cap D$. For example, let P and Q be computable probabilities on Ω, then $\frac{Q}{P}$ is a computable martingale in this sense. The following theorem shows martingale convergence theorem holds for individual random sequences. The proof is along the lines of the classical proof.

Theorem 3.1 (Doob) *Let $\{X_n\}$ be a computable submartingale. Assume that $\sup_n E(|X_n|) < \infty$. If $x^\infty \in \mathcal{R}^P$, then $\lim_{n\to\infty} X_n(x^\infty)$ exists and $\sup_n |X_n(x^\infty)| < \infty$.*

Let P be a computable probability on $X \times Y = \Omega^2$. Let P_X and P_Y be its marginal distributions on X and Y, respectively, i.e., $P_X(x) = P(x,\lambda)$ and $P_Y(y) = P(\lambda, y)$ for $x, y \in S$. Let

$$P(x|y) := \begin{cases} \frac{P(x,y)}{P_Y(y)}, & \text{if } P_Y(y) > 0 \\ 0, & \text{if } P_Y(y) = 0 \end{cases},$$

and

$$P(x|y^\infty) := \lim_{y\to y^\infty} P(x|y),$$

for $y^\infty \in \Omega$ if the right-hand side exists. It is known that $P(\cdot|y^\infty)$ is a probability measure on Ω for almost all y^∞ with respect to P_Y. Since $P(x|y)$ is a computable martingale for fixed x, by Theorem 3.1, we have a slightly stronger result as follows:

Theorem 3.2 *If $y^\infty \in \mathcal{R}^{P_Y}$, then $P(x|y^\infty)$ exists for all $x \in S$, and $P(\cdot|y^\infty)$ is a probability measure on $(\mathcal{B}_\Omega, \Omega)$.*

3.1 Fubini's theorem

Since $P(x,\cdot)$ is absolutely continuous relative to P_Y for a fixed x, by Radon-Nikodým theorem, we have

$$P(x,y) = \int_{\Delta(y)} P(x|y^\infty)dP_Y(y^\infty),$$

for $x,\ y \in S$. For a subset $A \subset X \times Y$ and $y^\infty \in Y$, set

$$A_{y^\infty} := \{x^\infty | (x^\infty, y^\infty) \in A\}.$$

A_{y^∞} is called y^∞-section of A. For example, $\mathcal{R}^P_{y^\infty} = \{x^\infty | (x^\infty, y^\infty) \in \mathcal{R}^P\}$. Since $P(\mathcal{R}^P) = 1$, we have

$$1 = P(\mathcal{R}^P) = \int_\Omega P(\mathcal{R}^P_{y^\infty} | y^\infty) dP_Y(y^\infty).$$

Therefore, $P(\mathcal{R}^P_{y^\infty} | y^\infty) = 1$ for almost all y^∞ with respect to P_Y. In the following, we present stronger results.

For simplicity, set $\tilde{U}_{y^\infty} := (\cap_n \tilde{U}_n)_{y^\infty}$. Since $\mathcal{R}^P = (\cap_n \tilde{U}_n)^c$, we have $\mathcal{R}^P_{y^\infty} = (\tilde{U}_{y^\infty})^c$.

Theorem 3.3 $\{y^\infty \mid P(\tilde{U}_{y^\infty} | y^\infty) > 0\} \subset (\mathcal{R}^{P_Y})^c$.

Corollary 3.1 *If $y^\infty \in \mathcal{R}^{P_Y}$, then $\sum_n P((\tilde{U}_n)_{y^\infty} | y^\infty) < \infty$.*

Lemma 3.1 $\mathcal{R}^P \subset \mathcal{R}^{P_X} \times \mathcal{R}^{P_Y}$.

Corollary 3.2 *$P(\mathcal{R}^P_{y^\infty} | y^\infty) = 1$ if $y^\infty \in \mathcal{R}^{P_Y}$. $\mathcal{R}^P_{y^\infty} = \emptyset$ if $y^\infty \notin \mathcal{R}^{P_Y}$.*

Corollary 3.3 $\mathcal{R}^{P_X} = \cup_{y^\infty \in \mathcal{R}^{P_Y}} \mathcal{R}^P_{y^\infty}$.

Note that except for trivial cases, $\mathcal{R}^P \neq \mathcal{R}^{P_X} \times \mathcal{R}^{P_Y}$. For example, let $\forall x, y,\ P(x, y) := P_X(x) P_X(y)$ for a computable probability P_X. Let $G := \{(x^\infty, x^\infty) | x^\infty \in \Omega\}$. If $P(G) = 0$ then we see that $G \cap \mathcal{R}^P = \emptyset$, and hence $\mathcal{R}^P \neq \mathcal{R}^{P_X} \times \mathcal{R}^{P_X}$.

For proofs, see [17]. In [20], Theorem 3.3 is shown for product probability, $P = P_X P_Y$.

4 Section of random set vs. relativized randomness

In this section we compare section of random set with relativized randomness. Let P_{y^∞} be a probability on Ω. We say that P_{y^∞} is computable relative to $y^\infty \in \Omega$ if there is a function $A^{y^\infty} : S \times \mathbb{N} \to \mathbb{Q}$ such that

$$\forall x \in S\ \forall k \in \mathbb{N},\ |A^{y^\infty}(x, k) - P_{y^\infty}(x)| < 1/k, \tag{3}$$

where A^{y^∞} is computable by a Turing machine with an auxiliary tape that contains y^∞.

Similarly, we say that a set $U^{y^\infty} \subset S$ is a r.e. set relative to y^∞ if U^{y^∞} is the range of a computable function relative to y^∞. Let P_{y^∞} be a computable probability relative to y^∞; then, we can define a relativized test U^{y^∞} of P_{y^∞}. Similarly to Theorem 2.1, we can show that a relativized universal test exists as follows:

Theorem 4.1 (relativized version of Martin-Löf theorem) *Let P_{y^∞} be a computable probability relative to y^∞ on Ω. Then, a universal test relative to y^∞ exists.*

Let $\{U_n^{y^\infty}\}$ be a relativized universal test with respect to P_{y^∞} and y^∞, and let

$$\mathcal{R}^{P_{y^\infty}, y^\infty} := (\cap_n \tilde{U}_n^{y^\infty})^c.$$

Note that the relativized universal test $\{U_n^{y^\infty}\}$ depends on P_{y^∞} and y^∞.

Recall that if $y^\infty \in \mathcal{R}^{P_Y}$, then the conditional probability $P(\cdot|y^\infty)$ exists (Theorem 3.2). By Corollary 3.1, we have

Theorem 4.2 *Let P be a computable probability on $X \times Y$ $(= \Omega^2)$ and P_Y be the marginal distribution on Y. If $P(\cdot|y^\infty)$ is computable relative to $y^\infty \in \mathcal{R}^{P_Y}$, then $\mathcal{R}^{P(\cdot|y^\infty), y^\infty} \subset \mathcal{R}^P_{y^\infty}$.*

In order to show the converse inclusion of the above theorem, we introduce a stronger notion of relative computability. Let A^{y^∞} be the relative computable function appeared in (3). In the course of the computation of $A^{y^\infty}(x,k)$, it uses only finite prefix of y^∞. Thus there is a partial computable function A such that

$$\forall x \in S \; \forall k \in \mathbb{N} \; \exists y \sqsubset y^\infty, \; A^{y^\infty}(x,k) = A(x,k,y), \tag{4}$$

and if $A(x,k,y)$ is defined then $A(x,k,y) = A(x,k,y')$ for all y' such that $y \sqsubset y'$.

Similarly, let U^{y^∞} be a relativized universal test of P_{y^∞}; then, there is a computable function B^{y^∞} relative to y^∞ and a partial computable function B such that

$$\forall n, \; U_n^{y^\infty} = \{x \in S | \exists i, \; B^{y^\infty}(i,n) = x\},$$

and

$$\forall i, n, \; \exists y \sqsubset y^\infty, \; B^{y^\infty}(i,n) = B(i,n,y). \tag{5}$$

If $B(i,n,y)$ is defined then $B(i,n,y) = B(i,n,y')$ for all y' such that $y \sqsubset y'$.

We say that the family $\{P_{y^\infty}\}_{y^\infty}$ is *uniformly computable in* $\mathcal{R}^{P_Y}$ if 1) P_{y^∞} is a computable probability relative to y^∞ for all $y^\infty \in \mathcal{R}^{P_Y}$ and 2) (4) holds for all $y^\infty \in \mathcal{R}^{P_Y}$, i.e., there is a partial computable function A such that

$$\forall y^\infty \in \mathcal{R}^{P_Y}\ \forall x \in S\ \forall k \in \mathbb{N}\ \exists y \sqsubset y^\infty,\ A^{y^\infty}(x,k) = A(x,k,y). \tag{6}$$

Theorem 4.3 *Let P be a computable probability on $X \times Y$ $(= \Omega^2)$ and P_Y be the marginal distribution on Y. If $\{P(\cdot|y^\infty)\}_{y^\infty}$ is uniformly computable in $\mathcal{R}^{P_Y}$, then $\mathcal{R}^P_{y^\infty} = \mathcal{R}^{P(\cdot|y^\infty),y^\infty}$ for $y^\infty \in \mathcal{R}^{P_Y}$.*

For proofs, see [17]. Note that section of random set is determined by global probability P and relativized randomness is determined locally by conditional probability.

5 Likelihood ratio test

Let P and Q be computable probabilities on Ω. Let

$$r(x) := \begin{cases} \frac{Q(x)}{P(x)}, & \text{if } P(x) > 0 \\ 0, & \text{if } P(x) = 0 \end{cases},$$

for $x \in S$. We see that r is a computable martingale. By the martingale convergence theorem for algorithmically random sequences, we have

Corollary 5.1 $\mathcal{R}^P \subset \{x^\infty | \lim_{x\to x^\infty} r(x) < \infty\}$.

Lemma 5.1 *Let P and Q be computable probabilities on Ω.*
1) : $\mathcal{R}^P \cap \mathcal{R}^Q = \mathcal{R}^P \cap \{x^\infty | 0 < \lim_{x\to x^\infty} r(x) < \infty\}$.
2) : $\mathcal{R}^P \cap (\mathcal{R}^Q)^c = \mathcal{R}^P \cap \{x^\infty | \lim_{x\to x^\infty} r(x) = 0\}$.

Proof) 1) If $x^\infty \in \mathcal{R}^P \cap \mathcal{R}^Q$, then a) by Corollary 5.1, $\lim_{x\to x^\infty} r(x) < \infty$ and $\lim_{x\to x^\infty} r^{-1}(x) < \infty$, and b) $P(x) > 0$ and $Q(x) > 0$ for $x \sqsubset x^\infty$; thus, $0 < \lim_{x\to x^\infty} r(x) < \infty$. Conversely, if $x^\infty \in \mathcal{R}^P \cap \{x^\infty | 0 < \lim_{x\to x^\infty} r(x) < \infty\}$, by Theorem 2.2, $\sup_{x\sqsubset x^\infty} -\log P(x) - Km(x) < \infty$ and $\sup_{x\sqsubset x^\infty} | -\log Q(x) + \log P(x)| < \infty$. Thus, $\sup_{x\in A(x^\infty)} -\log Q(x) - Km(x) < \infty$ and we have $x^\infty \in \mathcal{R}^Q$.
2) By 1, we have $\mathcal{R}^P \cap (\mathcal{R}^Q)^c = \mathcal{R}^P \cap (\mathcal{R}^P \cap \mathcal{R}^Q)^c = \mathcal{R}^P \cap (\{\lim r =$

$0\} \cup \{\lim r = \infty\}) = \mathcal{R}^P \cap \{\lim r = 0\}$, where the last equality follows from Corollary 5.1. ■

For other proof, see [3]. From the above lemma, we have the following:

Theorem 5.1 *Let P and Q be computable probabilities on Ω.*

$$\mathcal{R}^P \cap (\mathcal{R}^Q)^c = (\mathcal{R}^P \cup \mathcal{R}^Q) \cap \{x^\infty | \lim_{x \to x^\infty} r(x) = 0\}. \tag{7}$$

$$(\mathcal{R}^P)^c \cap \mathcal{R}^Q = (\mathcal{R}^P \cup \mathcal{R}^Q) \cap \{x^\infty | \lim_{x \to x^\infty} r(x) = \infty\}. \tag{8}$$

$$\mathcal{R}^P \cap \mathcal{R}^Q = (\mathcal{R}^P \cup \mathcal{R}^Q) \cap \{x^\infty | 0 < \lim_{x \to x^\infty} r(x) < \infty\}. \tag{9}$$

5.1 Absolute continuity and mutual singularity

By Lebesgue decomposition theorem, there exists $N \in \mathcal{F}_\infty$ such that $P(N) = 0$ and

$$\forall C \in \mathcal{F}_\infty,\ Q(C) = \int_C r(x^\infty) dP + Q(C \cap N). \tag{10}$$

We write (a) $P \perp Q$ if P and Q are mutually singular, i.e., there exist A and B such that $A \cap B = \emptyset$, $P(A) = 1$, and $Q(B) = 1$, and (b) $P \ll Q$ if P is absolutely continuous with respect to Q, i.e., $\forall C \in \mathcal{F}_\infty\ Q(C) = 0 \Rightarrow P(C) = 0$.

Remark 2 By (10), we have (a) $P \perp Q$ iff $P(\{\lim r = 0\}) = 1$, and (b) $P \ll Q$ iff $P(\{\lim r = 0\}) = 0$; for example, see [14].

The following theorem appeared in pp. 103 of [13] without proof.

Theorem 5.2 (Martin-Löf) *Let P and Q be computable probabilities on Ω. Then, $\mathcal{R}^P \cap \mathcal{R}^Q = \emptyset$ iff $P \perp Q$.*

Proof) Since $P(\mathcal{R}^P) = Q(\mathcal{R}^Q) = 1$, only if part follows. Conversely, assume that $P \perp Q$. Let $N := \{x^\infty | 0 < \liminf_{x \sqsubset x^\infty} r(x) \le \limsup_{x \sqsubset x^\infty} r(x) < \infty\}$. By Remark 2, we have $P(N) = Q(N) = 0$. Since $0 < \liminf_{x \sqsubset x^\infty} r(x) \Leftrightarrow 0 < \inf_{x \sqsubset x^\infty} r(x)$ and $\limsup_{x \sqsubset x^\infty} r(x) < \infty \Leftrightarrow \sup_{x \sqsubset x^\infty} r(x) < \infty$, we have

$$\begin{aligned} N &= \{x^\infty | 0 < \inf_{x \sqsubset x^\infty} r(x) \le \sup_{x \sqsubset x^\infty} r(x) < \infty\} \\ &= \cup_{a,b \in \mathbb{Q}, 0<a<b<\infty} \cap_{i=1}^\infty \tilde{N}_i^{a,b}, \end{aligned}$$

where $N_i^{a,b} = \{x | a \le r(y) \le b, \forall y \sqsubseteq x, |x| = i\}$. Since $P(N) = 0$, we have $\lim_i P(\tilde{N}_i^{a,b}) = 0$. Since $(N_i^{a,b})^c \cap \{x | P(x) > 0\}$ is a r.e. set, we can approximate $P(\tilde{N}_i^{a,b})$ from above, and there is a computable function $\alpha(n)$ such that $P(\tilde{N}_{\alpha(n)}^{a,b}) < 2^{-n}$. Thus, $\tilde{N}_{\alpha(n)}^{a,b}$ is a test of P, and hence, $N \subset (\mathcal{R}^P)^c$. Similarly, we have $N \subset (\mathcal{R}^Q)^c$. By (9), we have $\mathcal{R}^P \cap \mathcal{R}^Q = \emptyset$. ∎

Lemma 5.2 *Let P and Q be computable probabilities on Ω. Then,* $\mathcal{R}^P \subset \mathcal{R}^Q \Rightarrow P \ll Q$.

There is a counter example for the converse implication of the above lemma, see [3].

5.2 Countable model class

In the following discussion, let $\{P_n\}_{n\in\mathbb{N}}$ be a family of computable probabilities on Ω; more precisely, we assume that there is a computable function $A : \mathbb{N} \times S \times \mathbb{N} \to \mathbb{Q}$ such that $|A(n, x, k) - P_n(x)| < 1/k$ for all $n, k \in \mathbb{N}$ and $x \in S$. Note that we cannot set $\{P_n\}_{n\in\mathbb{N}}$ as the entire family of computable probabilities on Ω since it is not a r.e. set. Let α be a computable positive probability on $\mathbb{N}$, i.e., $\forall n\, \alpha(n) > 0$ and $\sum_n \alpha(n) = 1$. Then, set $P := \sum_n \alpha(n) P_n$. We see that P is a computable probability. The following lemma shows that the set of random sequences of a discrete mixture of computable probabilities are the union of their random sets.

Lemma 5.3 $\mathcal{R}^P = \cup_n \mathcal{R}^{P_n}$.

Let β be a computable probability on $\mathbb{N}$ such that 1) $\beta(n) > 0$ if $n \neq n^*$ and $\beta(n^*) = 0$, and 2) $\sum_n \beta(n) = 1$. Then, set

$$P^- := \sum_n \beta(n) P_n.$$

We see that P^- is a computable probability. By Theorem 5.1 and Lemma 5.3, we have

Lemma 5.4
$\mathcal{R}^{P_{n^*}} \cap_{n\neq n^*} (\mathcal{R}^{P_n})^c = (\cup_n \mathcal{R}^{P_n}) \cap \{x^\infty | \lim_{x\to x^\infty} P^-(x)/P_{n^*}(x) = 0\}$.

Now let

$$\hat{n}(x) := \arg\max_n \alpha(n) P_n(x).$$

Then we can show that

$$\lim_{x \to x^\infty} P^-(x)/P_{n^*}(x) = 0 \Rightarrow \lim_{x \to x^\infty} \hat{n}(x) = n^*.$$

Thus we have

$$\mathcal{R}^{P_{n^*}} \cap_{n \neq n^*} (\mathcal{R}^{P_n})^c \subset \{x^\infty | \lim_{x \to x^\infty} \hat{n}(x) = n^*\},$$

which shows that if x^∞ is random with respect to $\mathcal{R}^{P_{n^*}}$ and it is not random with respect to other models then $\hat{n}$ classifies its model. Estimation of models by $\hat{n}$ is called MDL model selection, for more details, see [1, 2]. Note that by Theorem 5.2, if $\{P_n\}$ are mutually singular, then $\mathcal{R}^{P_{n^*}} \cap_{n \neq n^*} (\mathcal{R}^{P_n})^c = \mathcal{R}^{P_{n^*}}$, and by Lemma 5.2, if $P_{n^*} \not\ll P^-$, then $\mathcal{R}^{P_{n^*}} \cap_{n \neq n^*} (\mathcal{R}^{P_n})^c \neq \emptyset$.

6 Decomposition of complexity

It is known that

$$\sup_{x,y \in S} |K(x,y) - K(x|y, K(y)) - K(y)| < \infty, \tag{11}$$

where K is the prefix Kolmogorov complexity [5, 8]. If we eliminate $K(y)$ from $K(x|y, K(y))$ in (11), then it is not asymptotically bounded, i.e.,

$$\sup_{x,y \in S} |K(x,y) - K(x|y) - K(y)| = \infty.$$

For more details, see [8]. Since $|Km(\bar{x}, \bar{y}) - K(x,y)| = O(1)$, $|Km(\bar{x}) - K(x)| = O(1)$, and $|Km(\bar{x}|\bar{y}) - K(x|y)| = O(1)$ (recall that $\overline{x} = 1^{|x|}0x$), we have

$$\sup_{x,y \in S} |Km(x,y) - Km(x|y) - Km(y)| = \infty. \tag{12}$$

The above equation shows that there is a sequence of strings such that left-hand side of the above equation is unbounded. However, if we restrict strings to prefixes of random sequences x^∞, y^∞ with respect to some computable probability, then we can show that (12) is bounded for a sufficiently large prefix of (x^∞, y^∞) under a condition (see Theorem 6.1 below).

Let $Km(x|y^\infty) := \lim_{y \to y^\infty} Km(x|y)$. Recall that $Km(x|y)$ is decreasing as $y \to y^\infty$. Then set

$$\alpha(x, y^\infty, c) := \inf\{y | \forall y',\ y \sqsubseteq y' \sqsubset y^\infty\ Km(x|y') - Km(x|y^\infty) \leq c\},$$

for $c \geq 0$. Since Km always takes an integer value, we see that $\alpha(x, y^\infty, c)$ always takes a finite prefix y of y^∞ for $c \geq 0$. Roughly speaking, if c is small then α has almost the same information that y^∞ has regarding x. For example, if $\alpha = \lambda$ and $c = 0$, then y^∞ contains no information about x.

Next, let

$$\beta(x, y^\infty, c) := \inf\{y | \forall y', \, y \sqsubseteq y' \sqsubset y^\infty, \, |\log P(x|y^\infty) - \log P(x|y)| \leq c\},$$

for $y^\infty \in \mathcal{R}^{P_Y}, c > 0$. Since $P(x|y) \to P(x|y^\infty)$ for $y^\infty \in \mathcal{R}^{P_Y}$, $\beta(x, y^\infty, c)$ takes a finite value for all x and $0 < c$. Intuitively, β is a convergence rate of $P(x|y)$. For example, if $P(x|y) = P(x)$, then $\beta = \lambda$.

Since $\alpha(x, y^\infty, c)$ and $\beta(x, y^\infty, c)$ are comparable, let

$$\gamma(x, y^\infty, c) := \sup\{\alpha(x, y^\infty, c), \beta(x, y^\infty, c)\}.$$

We say that γ is (x^∞, y^∞)-*recursively increasing* if there is a monotonically increasing recursive function $g : \mathbb{N} \to \mathbb{N}$ such that $\exists c \forall x \sqsubset x^\infty$, $|\gamma(x, y^\infty, c)| \leq g(|x|)$. g is called *recursive upper function.*

Theorem 6.1 *Let P be a computable probability on Ω^2. Let $(x^\infty, y^\infty) \in \mathcal{R}^P$. Assume that $P(\cdot|y^\infty)$ is computable relative to $y^\infty \in \mathcal{R}^{P_Y}$ and γ is (x^∞, y^∞)-recursively increasing. Let g be a recursive upper function. Then for the partition function $f(n) = (n, g(n))$, we have*

$$\sup_{(x,y) \in \mathcal{A}^f(x^\infty, y^\infty)} |Km(x, y) - Km(x|y) - Km(y)| < \infty. \tag{13}$$

Proof) Assume that $(x^\infty, y^\infty) \in \mathcal{R}^P$. By Corollary 2.1, we have

$$\sup_{(x,y) \in \mathcal{A}^f(x^\infty, y^\infty)} |\log P(x, y) + Km(x, y)| < \infty, \tag{14}$$

and

$$\sup_{y \sqsubset y^\infty} |\log P_Y(y) + Km(y)| < \infty, \tag{15}$$

where (15) follows from that $(x^\infty, y^\infty) \in \mathcal{R}^P \Rightarrow y^\infty \in \mathcal{R}^{P_Y}$.

Since $\exists c, \forall x, y, Km(x, y) \leq Km(x|y) + Km(y) + c$ (Proposition 2.1 3), we have

$$\sup_{(x,y) \in \mathcal{A}^f(x^\infty, y^\infty)} -\log P(x|y) - Km(x|y) < \infty. \tag{16}$$

Since $(x, y) \in \mathcal{A}^f(x^\infty, y^\infty)$ implies $\gamma(x, y^\infty, c) \sqsubseteq y$, we have

$$
\begin{aligned}
-\log P(x|y) &\geq -\log P(x|y^\infty) - c & (17)\\
&\geq Km(x|y^\infty) - c - c_1 & (18)\\
&= Km(x|y) - 2c - c_1. & (19)
\end{aligned}
$$

Here, (17) follows from $\beta(x, y^\infty, c) \sqsubseteq \gamma(x, y^\infty, c)$, (18) follows from that $P(\cdot|y^\infty)$ is a relative computable probability on Ω, where c_1 is a constant independent from x, and (19) follows from $\alpha(x, y^\infty, c) \sqsubseteq \gamma(x, y^\infty, c)$. Thus we have

$$\sup_{(x,y)\in\mathcal{A}^f(x^\infty,y^\infty)} |\log P(x|y) + Km(x|y)| < \infty. \tag{20}$$

By (14), (15), and (20), we have the theorem. ■

6.1 Relativized randomness

Next we compare a pair of randomness with relativized randomness. If $P(\cdot|y^\infty)$ is computable relative to y^∞, let $\mathcal{R}^{P(\cdot|y^\infty), y^\infty}$ be the set of random sequences with respect to $P(\cdot|y^\infty)$. Then we have the relativized version of Levin-Schnorr theorem.

Theorem 6.2 (relativized version of Levin-Schnorr thereom) *Let $P(\cdot|y^\infty)$ be a computable probability relative to y^∞ on Ω. Then,* $\mathcal{R}^{P(\cdot|y^\infty), y^\infty} = \{x^\infty | \sup_{x \sqsubset x^\infty} |\log P(x|y^\infty) + Km(x|y^\infty)| < \infty\}$.

The following corollary is shown in Theorem 4.3 under the uniform computability assumption. We show the same equivalence under the assumption that γ is (x^∞, y^∞)-recursively increasing.

Corollary 6.1 *Let P be a computable probability on Ω^2. Assume that $P(\cdot|y^\infty)$ is computable relative to $y^\infty \in \mathcal{R}^{P_Y}$ and γ is (x^∞, y^∞)-recursively increasing. Then we have $(x^\infty, y^\infty) \in \mathcal{R}^P$ iff $x^\infty \in \mathcal{R}^{P(\cdot|y^\infty), y^\infty}$.*

Proof) Let $f(n) = (n, g(n))$, where g is a recursive upper function. Since $(x, y) \in \mathcal{A}^f(x^\infty, y^\infty)$ implies $\gamma(x, y^\infty, c) \sqsubseteq y$, we have

$$\sup_{(x,y)\in\mathcal{A}^f(x^\infty,y^\infty)} |\log P(x|y) - \log P(x|y^\infty)| < \infty,$$

and

$$\sup_{(x,y)\in\mathcal{A}^f(x^\infty,y^\infty)} |Km(x|y) - Km(x|y^\infty)| < \infty.$$

Assume that $(x^\infty, y^\infty) \in \mathcal{R}^P$. From (20), we have

$$\sup_{x \sqsubset x^\infty} |\log P(x|y^\infty) + Km(x|y^\infty)| < \infty. \tag{21}$$

From Theorem 6.2, we have the only if part. The converse implication follows from Corollary 3.1. ■

6.2 Independence

Let

$$\alpha(x^\infty, y^\infty, c) := \sup_{x \sqsubset x^\infty} \alpha(x, y^\infty, c),$$

and

$$\beta(x^\infty, y^\infty, c) := \sup_{x \sqsubset x^\infty} \beta(x, y^\infty, c).$$

We may say that if $\exists\, 0 \leq c < \infty,\ |\alpha(x^\infty, y^\infty, c)| < \infty$, then x^∞ and y^∞ are algorithmically independent, and if $\exists\, 0 < c < \infty,\ |\beta(x^\infty, y^\infty, c)| < \infty$, then x^∞ and y^∞ are stochastically independent. In fact, we have

Theorem 6.3 *Let P be a computable probability on Ω^2 and $(x^\infty, y^\infty) \in \mathcal{R}^P$.*
A: The following statements (1), (2), and (3) are equivalent:
(1) $\exists\, 0 < c < \infty,\ |\beta(x^\infty, y^\infty, c)| < \infty$.
(2) $(x^\infty, y^\infty) \in \mathcal{R}^Q$, where Q is a computable probability on Ω^2 defined by $Q(x, y) := P_X(x)P_Y(y)$ for all $x, y \in S$.
(3) For any $\mathcal{A} \subset S^2$ that satisfies Condition 1 and 2,
$\sup_{(x,y) \in \mathcal{A}(x^\infty, y^\infty)} |Km(x, y) - Km(x) - Km(y)| < \infty$.
B: Assume that $x^\infty \in \mathcal{R}^{P(\cdot|y^\infty), y^\infty}$ and there is a monotonically increasing recursive function $g : \mathbb{N} \to \mathbb{N}$ such that $\exists c \forall x \sqsubset x^\infty,\ |\alpha(x, y^\infty, c)| \leq g(|x|)$. The statements (1), (2), and (3) are equivalent to
(4) $\exists\, 0 < c < \infty,\ |\alpha(x^\infty, y^\infty, c)| < \infty$.

Proof) A: (1) $\Rightarrow$ (2): First, we show that

$$0 < \lim_{(x,y) \to (x^\infty, y^\infty)} \frac{P_X(x)P_Y(y)}{P(x, y)} < \infty. \tag{22}$$

Since $(x^\infty, y^\infty) \in \mathcal{R}^P$, by Theorem 5.1 (it is easy to extend the theorem for computable probability on $\Omega \times \Omega$), $\lim_{(x,y) \to (x^\infty, y^\infty)} \frac{P_X(x)P_Y(x)}{P(x,y)}$ exists and is finite. Thus, we have to show that it is positive. For simplicity, let $\beta^\infty := \beta(x^\infty, y^\infty, c)$. We have

$$\lim_{x\to x^\infty, y\to y^\infty} \frac{P_X(x)P_Y(y)}{P(x,y)} = \lim_{x\to x^\infty} \frac{P_X(x)}{P(x|y^\infty)} \geq 2^{-c} \lim_{x\to x^\infty} \frac{P_X(x)}{P(x|\beta(x,y^\infty,c))}$$
$$\geq 2^{-2c} \lim_{x\to x^\infty} \frac{P_X(x)}{P(x|\beta^\infty)},$$

where the first equality follows from that $P(x|y^\infty)$ exists for $y^\infty \in \mathcal{R}^{P_Y}$. On the other hand, since $|\beta^\infty| < \infty$, we have $x^\infty \in \mathcal{R}^{P(\cdot|\beta^\infty)}$ and $x^\infty \in \mathcal{R}^{P_X}$. Therefore, by Theorem 5.1, $\lim_{x\to x^\infty} \frac{P_X(x)}{P(x|\beta^\infty)} > 0$, and (22) holds. From Theorem 5.1, we have the statement (2).
(2) $\Rightarrow$ (3): Let $\mathcal{A}$ be a partition that satisfies Condition 1 and 2. By Theorem 2.3, we have

$$\sup_{(x,y)\in\mathcal{A}(x^\infty,y^\infty)} -\log P_X(x) - \log P_Y(y) - Km(x,y) < \infty.$$

Since 1) $x^\infty \in \mathcal{R}^{P_X}$ and $y^\infty \in \mathcal{R}^{P_Y}$, and 2) P_X and P_Y are computable probabilities on Ω, we have $\sup_{x\sqsubset x^\infty} |Km(x)+\log P_X(x)| < \infty$, and $\sup_{y\sqsubset y^\infty} |Km(y)+\log P_Y(y)| < \infty$. On the other hand,

$$\exists c > 0\, \forall x,y \in S\;\; Km(x,y) \leq Km(x|y) + Km(y) + c \leq Km(x) + Km(y) + c. \tag{23}$$

Thus, we have statement (3).
(3) $\Rightarrow$ (1): Let $\mathcal{A} = \{(x,y) | |x| = |y|\}$. Since $(x^\infty, y^\infty) \in \mathcal{R}^P$, $x^\infty \in \mathcal{R}^{P_X}$, and $y^\infty \in \mathcal{R}^{P_Y}$, by Corollary 2.1, we have $\sup_{(x,y)\in\mathcal{A}(x^\infty,y^\infty)} |Km(x,y) + \log P(x,y)| < \infty$, $\sup_{x\sqsubset x^\infty} |Km(x)+\log P_X(x)| < \infty$, and $\sup_{y\sqsubset y^\infty} |Km(y)+\log P_Y(y)| < \infty$. Thus we have

$$\exists c \forall (x,y) \in \mathcal{A}(x^\infty, y^\infty),\; 2^{-c} \leq \frac{P_X(x)P_Y(y)}{P(x,y)} \leq 2^c. \tag{24}$$

Since $(x^\infty, y^\infty) \in \mathcal{R}^P$, by Theorem 5.1, $\lim_{(x,y)\to(x^\infty,y^\infty)} \frac{P_X(x)P_Y(x)}{P(x,y)}$ exists and is finite. In particular, from (24), we have

$$\exists c\, \exists (x', y') \forall (x', y') \sqsubseteq (x,y) \sqsubset (x^\infty, y^\infty), 2^{-c} \leq \frac{P_X(x)P_Y(y)}{P(x,y)} \leq 2^c,$$

i.e., $\exists c \forall (x,y) \sqsubset (x^\infty, y^\infty)\; |\log P_X(x) - \log P(x|y)| < c$, which shows the statement (1).

B: (3) $\Rightarrow$ (4): Let $f = (n, g(n))$. By (23) and statement (3), we have $\sup_{(x,y)\in\mathcal{A}^f(x^\infty,y^\infty)} Km(x) - Km(x|y) < \infty$. Since $(x,y) \in \mathcal{A}^f(x^\infty, y^\infty)$ implies $\alpha(x, y^\infty, c) \sqsubset y$, we have $\sup_{(x,y)\in\mathcal{A}^f(x^\infty,y^\infty)} Km(x|y) - Km(x|y^\infty) \le c$. Since $Km(x) \ge Km(x|y) \ge Km(x|y^\infty)$, we have $\exists c,\ \alpha(x^\infty, y^\infty, c) = \lambda$.
(4) $\Rightarrow$ (2): By Theorem 5.1, it is enough to show (22). Since $(x^\infty, y^\infty) \in \mathcal{R}^P$, the limit exists and is finite. Thus it is enough to show

$$0 < \lim_{(x,y)\to(x^\infty,y^\infty)} \frac{P_X(x)P_Y(y)}{P(x,y)} = \lim_{x\to x^\infty} \frac{P_X(x)}{P(x|y^\infty)}. \tag{25}$$

Since $x^\infty \in \mathcal{R}^{P_X}$ and $x^\infty \in \mathcal{R}^{P(\cdot|y^\infty), y^\infty}$, from Theorem 6.2, we have $\sup_{x\sqsubset x^\infty} |Km(x) + \log P_X(x)| < \infty$ and $\sup_{x\sqsubset x^\infty} |\log P(x|y^\infty) + Km(x|y^\infty)| < \infty$. Hence from the statement (4), we have (25). ■

7 Bayesian statistics

Let P be a computable probability on $X \times Y$ and P_X, P_Y be its marginal distributions as before. In Bayesian statistical terminology, if X is a sample space, then P_X is called mixture distribution, and if Y is a parameter space, then P_Y is called prior distribution. In this section, we show that section of random set satisfies many theorem of Bayesian statistics and it is natural as a definition of random set with respect to conditional probability from Bayesian statistical point of view.

7.1 Optimality of Bayes code

A universal coding obtained by applying arithmetic coding to P_X is called *Bayes coding*. It is known that Bayes coding is optimal for $P(\cdot|y^\infty)$-almost all samples for almost all y^∞ with respect to P_Y, see [2]. We have a slightly stronger result.

Corollary 7.1 *The following three statements are equivalent:*
(1) $x^\infty \in \mathcal{R}^{P_X}$.
(2) $\sup_{x\sqsubset x^\infty} -\log P_X(x) - Km(x) < \infty$.
(3) $\exists y^\infty \in \mathcal{R}^{P_Y},\ x^\infty \in \mathcal{R}^P_{y^\infty}$.

Proof) (1) $\Leftrightarrow$ (2) follows from Theorem 2.2. (1) $\Leftrightarrow$ (3) follows from Corollary 3.3. ■

7.2 Consistency of posterior distribution

In this section, we show a consistency of posterior distribution for algorithmically random sequences. We see that the classification of random sets by likelihood ratio test (see Section 5) plays an important role in this section.

Theorem 7.1 *Let P be a computable probability on $X \times Y$, where $X = Y = \Omega$. Assume that $m(y) > 0$ for all $y \in S$. The following six statements are equivalent:*
(1) $P(\cdot|y) \perp P(\cdot|z)$ if $\Delta(y) \cap \Delta(z) = \emptyset$ for $y, z \in S$.
(2) $\mathcal{R}^{P(\cdot|y)} \cap \mathcal{R}^{P(\cdot|z)} = \emptyset$ if $\Delta(y) \cap \Delta(z) = \emptyset$ for $y, z \in S$.
(3) $P_{Y|X}(\cdot|x)$ converges weakly to I_{y^∞} as $x \to x^\infty$ for $(x^\infty, y^\infty) \in \mathcal{R}^P$, where I_{y^∞} is the probability that has probability of 1 at y^∞.
(4) $\mathcal{R}^P_{y^\infty} \cap \mathcal{R}^P_{z^\infty} = \emptyset$ if $y^\infty \neq z^\infty$.
(5) There exists a surjective function $f : \mathcal{R}^{P_X} \to \mathcal{R}^{P_Y}$ such that $f(x^\infty) = y^\infty$ for $(x^\infty, y^\infty) \in \mathcal{R}^P$.
(6) There exists $f : X \to Y$ and $Y' \subset Y$ such that $m(Y') = 1$ and $f = y^\infty$, $P(\cdot|y^\infty) - a.s.$ for $y^\infty \in Y'$.

Proof) (1) $\Leftrightarrow$ (2) follows from Theorem 5.2.
(2) $\Rightarrow$ (3) : If $(x^\infty, y^\infty) \in \mathcal{R}^P$, then $x^\infty \in \mathcal{R}^{P(\cdot|y)}$ for $y^\infty \in \Delta(y)$, see Lemma 3.1. By assumption if $\Delta(y) \cap \Delta(z) = \emptyset$, then $x^\infty \notin \mathcal{R}^{P(\cdot|z)}$. By Theorem 5.1, we have $\lim_{x \to x^\infty} P(x|z)/P(x|y) = 0$, and

$$\begin{aligned} \lim_{x \to x^\infty} \frac{P(x|z)}{P(x|y)} = 0 \quad &\Leftrightarrow \quad \lim_{x \to x^\infty} \frac{P(x,z)}{P(x,y)} = 0 \\ &\Leftrightarrow \quad \lim_{x \to x^\infty} \frac{P_{Y|X}(z|x)}{P_{Y|X}(y|x)} = 0. \end{aligned} \tag{26}$$

Since (26) holds for an arbitrary $\Delta(y)$ and $\Delta(z)$ such that $\Delta(y) \cap \Delta(z) = \emptyset$ and $y^\infty \in \Delta(y)$, we see that the posterior distribution $P_{Y|X}(\cdot|x)$ converges weakly to I_{y^∞}.
(3) $\Rightarrow$ (4) : obvious.
(4) $\Rightarrow$ (5) : Since $\mathcal{R}^P_{y^\infty} \cap \mathcal{R}^P_{z^\infty} = \emptyset$ for $y^\infty \neq z^\infty$, we can define a function $f : X \to Y$ such that $f(x^\infty) = y^\infty$ for $x^\infty \in \mathcal{R}^P_{y^\infty}$. Since by Corollary 3.3, $\mathcal{R}^{P_X} = \{x^\infty | (x^\infty, y^\infty) \in \mathcal{R}^P\}$ and $\mathcal{R}^{P_Y} = \{y^\infty | (x^\infty, y^\infty) \in \mathcal{R}^P\}$, and we have (5).
(5) $\Rightarrow$ (6) : By theorem 3.3, we have (6).

(6) $\Rightarrow$ (1) : Let $A_{y^\infty} := \{x^\infty | f(x^\infty) = y^\infty\}$. Then, $A_{y^\infty} \cap A_{z^\infty} = \emptyset$ for $y^\infty \neq z^\infty$ and $P(A_{y^\infty}|y^\infty) = 1$ for $y^\infty \in Y'$. Thus,
$(\cup_{y^\infty \in \Delta(y)} A_{y^\infty}) \cap (\cup_{y^\infty \in \Delta(z)} A_{y^\infty}) = \emptyset$ for $\Delta(y) \cap \Delta(z) = \emptyset$ and
$P(\cup_{y^\infty \in \Delta(y)} A_{y^\infty}|y) = P(\cup_{y^\infty \in \Delta(z)} A_{y^\infty}|z) = 1$, which shows (1). ∎

Example 1 Bernoulli model: Let $f(x^\infty) := \lim_n(\sum_{i=1}^n x_i)/n$ for $x^\infty = x_1 x_2 \cdots$. By the law of large numbers, (6) (and all the statements) are satisfied.

7.3 Algorithmically best estimator

Theorem 7.2 *Let P be a computable probability on $X \times Y$, where $X = Y = \Omega$. Let $f : \mathbb{N} \to \mathbb{N}$ be an unbounded increasing recursive function. Let $y^\infty \in Y$, and let $y_{f(n)}$ be a prefix of y^∞ of length $f(n)$*
(a) If $\lim_{x \to x^\infty} -\log P(y_{f(|x|)}|x) < \infty$, then there is a computable function ρ such that $y_{f(|x|)} = \rho(x)$ for infinitely many prefix x of x^∞.
(b) If $(x^\infty, y^\infty) \in \mathcal{R}^P$ and $\lim_{x \to x^\infty} -\log P(y_{f(|x|)}|x) = \infty$, then there is no computable monotone function ρ such that $\forall x \sqsubset x^\infty$, $y_{f(|x|)} \sqsubseteq \rho(x)$.

Proof) (a) By applying Shannon-Fano-Elias coding to $P(\cdot|x)$ on the finite partition $\{y||y| = f(|x|)\}$, we can construct a computable function g and a program $p \in S$ such that $g(p, x) = y$ and $|p| = \lceil -\log P(y|x) \rceil + 1$. Here, g need not be a monotone function. Since $|p| < \infty$ as $x \to x^\infty$, there is a p_0 such that $g(p_0, x) = y$ for infinitely many prefix x of x^∞. Thus, $\rho(x) := g(p_0, x)$ satisfies (a).
(b) By considering the partition function $f'(n) = (n, f(n))$ in (16), if $(x^\infty, y^\infty) \in \mathcal{R}^P$ and $\lim_{x \to x^\infty} -\log P(y_{f(|x|)}|x) = \infty$, then we have $\lim_{x \to x^\infty} Km(y_{f(|x|)}|x) = \infty$. Note that in order to show (16), the condition about γ is not necessary. Now assume that there is a computable monotone function ρ such that $\forall x \sqsubset x^\infty$ $y_{f(|x|)} \sqsubseteq \rho(x)$. Then, $\lim_{x \to x^\infty} Km(y_{f(|x|)}|x) < \infty$, which is a contradiction. ∎

By definition, we have

$$-\log P(y|x) = -\log \int_{\Delta(y)} P(x|y^\infty) dP_Y(y^\infty) + \log \int_Y P(x|y^\infty) dP_Y(y^\infty). \tag{27}$$

Let P_Y be a Lebesgue absolutely continuous measure. Let $\hat{y}$ be the maximum likelihood estimator. By using Laplace approximation with suitable conditions, if $\hat{y} \in \Delta(y)$ and $f(|x|) \approx \frac{1}{2}\log|x|$, then the right-hand-side of (27) is asymptotically bounded, for example see [1], and we have $\lim_{x \to x^\infty} -\log P(y|x) < \infty$, where $|y| = f(|x|)$. Thus, by Theorem 7.2 (a), we can compute initial $\lceil \frac{1}{2}\log|x| \rceil$-bits of y^∞ from x infinitely many times, which is an algorithmic version of a well known result in statistics: $|y^\infty - \hat{y}| = O(1/\sqrt{n})$.

Let $f(\cdot)$ be a large order function such that $\lim_{x \to x^\infty} -\log P(y|x) = \infty$ for $|y| = f(|x|)$; for example, set $f(|x|) = \lceil \log|x| \rceil$. By Theorem 7.2 (b), there is no monotone computable function that computes initial $f(|x|)$-bits of y^∞ for all $x \sqsubset x^\infty$. If such a function exists, then y^∞ is not random with respect to P_Y and the Lebesgue measure of such parameters is 0. On the other hand, it is known that the set of parameters that are estimated within $o(1/\sqrt{n})$ accuracy has Lebesgue measure 0 [4].

Theorem 7.2 shows a relation between the redundancy of universal coding and parameter estimation; as in [18], if we set P_Y to be a singular prior, $\lim_{x \to x^\infty} -\log P(y_f|x) < \infty$ for a large order f. In such a case we have a super-efficient estimator.

Acknowledgement

The author thanks Prof. Teturo Kamae (Matsuyama Univ.), Prof. Akio Fujiwara, Masahiro Nakamura (Osaka Univ.) and Prof. Alexander Shen (Marseilles, France) for discussions and comments.

References

[1] A. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling. *IEEE Trans. Inform. Theory*, 44(6):2743–2760, 1998.

[2] A. R. Barron. *Logically smooth density estimation.* PhD thesis, Stanford Univ., 1985.

[3] L. Bienvenu and W. Merkle. Effective randomness for computable probability measures. *Electron. Notes Theor. Comput. Sci.*, 167:117–130, 2007.

[4] L. Le Cam. On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications in Statistics*, 1:277–330, 1953.

[5] G. J. Chaitin. A theory of program size formally identical to information theory. *J. ACM*, 22:329–340, 1975.

[6] A. Chernov and M. Hutter. Monotone conditional complexity bounds on future prediction errors. In *ALT 2005*, volume 3734 of *LNAI*, pages 414–428. Springer, 2005.

[7] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, Hoboken, New Jersey, second edition, 2006.

[8] P. Gács. On the symmetry of algorithmic information. *Soviet Math. Dokl.*, 15(5):1477–1480, 1974.

[9] P. Gács. Uniform test of algorithmic randomness over a general space. *Theoret. Comput. Sci.*, 341:91–137, 2005.

[10] L. A. Levin. On the notion of a random sequence. *Soviet. Math. Dokl.*, 14(5):1413–1416, 1973.

[11] M. Li and P. Vitányi. *An introduction to Kolmogorov complexity and Its applications.* Springer, New York, second edition, 1997.

[12] P. Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–609, 1966.

[13] P. Martin-Löf. *Notes on constructive mathematics.* Almqvist & Wiksell, Stockholm, 1968.

[14] J. Neveu. *Discrete-Parameter Martingales.* North-Holland, Amsterdam, 1975.

[15] C. P. Schnorr. Process complexity and effective random tests. *J. Comp. Sys. Sci.*, 7:376–388, 1973.

[16] C. P. Schnorr. A survey of the theory of random sequences. In Butts and Hintikka, editors, *Basic problems in Methodology and Linguistics*, pages 193–211. Reidel, Dordrecht, 1977.

[17] H. Takahashi. On a definition of random sequences with respect to conditional probability. To appear in Information and Computation.

[18] H. Takahashi. Redundancy of universal coding, Kolmogorov complexity, and Hausdorff dimension. *IEEE Trans. Inform. Theory*, 50(11):2727–2736, 2004.

[19] V. A. Uspenskii, A. L. Semenov, and A. Kh. Shen. Can an individual sequence of zeros and ones be random? *Russian Math. Surveys*, 45(1):121–189, 1990.

[20] M. van Lambalgen. *Random sequences.* PhD thesis, Universiteit van Amsterdam, 1987.